17-JUL-75 17:34:41-PDT,3377;000000000000
Date: 17 JUL 1975 1734-PDT
From: RINDFLEISCH
Subject: COMMENTS ON WIPKE PROPOSAL
To: LEDERBERG, LEVINTHAL, FEIGENBAUM, BUCHANAN, SMITH

The following are some comments on the Wipke proposal from the view
point of system loading. I have only cursorily looked over his
communications estimates and have not had time to verify his numbers
with the telephone company here.

There are two major areas of comment; 1) the realism of his estimates
and 2) the relation between his proposed project and current system

use. His estimates of system loading include:

        CPU time   22 - 65 hours/month
        Connect time   430 hours/month
        File space   8000 pages

These are to be compared to the following data for currently active
projects (averages over April, May, and June):

	CPU Hrs.	Conn Hrs.	File Pgs
DIALOG	4.63	164	5100
Rutgers	12.15	488	6210
DENDRAL	51.68	1375	19020
MYCIN	20.16	484	4500
Protein Model	5.56	167	1220

There is a great deal of uncertainty in his estimates of CPU time
(factor of 3). It seems clear that the high end cannot be correct,
for example as compared to the DENDRAL usage. This is borne out in
some conversations with Peter Friedland about the relative demands
of Wipke's programs and say CONGEN. Wipke's code is all in FORTRAN
as opposed to LISP and should place considerably less demand. I
think there are a lot of uncertainties is Wipke's estimates for
legitimate reasons (he is coming from a small KA-10 [64K] under an
ancient DEC system to a large core KI-TENEX). Friedland felt that much
time was spent in disk access for overlays. I am trying to pin some
of these numbers down better. I am trying to reach Wipke to propose
that comparative runs be made with the "Model Builder" on his machine
and with the version running here - this may provide a scale for the
translation. It is not clear this can be done before Wipke leaves
Princeton.

If we take the lower estimate of CPU usage, his project would be
about of the scale of MYCIN. His statement that "not all of this time
would be needed during prime shift" and that he "would be willing
to assist the community in the solution to loading problems" suggest
that his work could be accommodated at this lower level of usage.

The terminal access time does not seem grossly out of reason. As I
said above, I can't confirm his estimates of line charges.

His estimates of file requirements are high; at least as I read the
rationale for his numbers. The 8000 page estimate seems to be based
on multiple copies of SECS (about 5 or 6 copies at 1200 pages each).
It seems like poor software control to have so many copies of sources,
rel files and binaries lying around! Data files I could understand
but these are not mentioned. Thus I think his requirements for disk
space may reflect the fact that he was on a one man machine with lots
of file space and he expanded to fill it. It is hard to estimate the
real need for programs and data - perhaps 2 versions plus data at
around 3000-4000 pages would be enough. If his file needs are indeed
as large as he says (40% of a disk drive), then we will need to ask
his grant proposal to include funds to augment our file system
appropriately.

; <LEVINTHAL>WIPKE.;1   MON 21-JUL-75 9:40AM

I will pass on other data as it becomes available.

Tom R.